# RESIDUAL PLOTS

5 assumptions in Residual plot – i) Linear ii) Constant Variance iii) Normality (Residuals) iv) Random (Independent) v) Residual average zero

1) Linear → curve, bend, flow

2) Constant Variance → spreading.

3) Normality (Residuals) → outliers, density on one side.

4) Random (Independent) → Bias in the data. ie. selection bias.

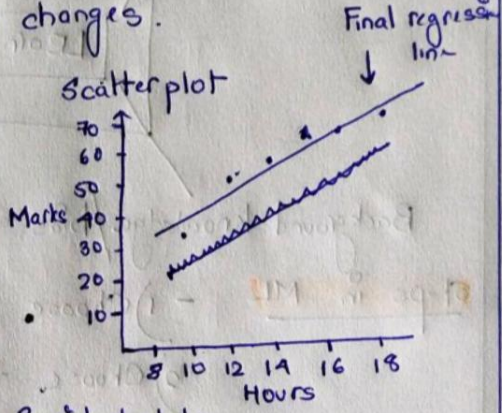
5) Residual average zero → How far you are from regression line?

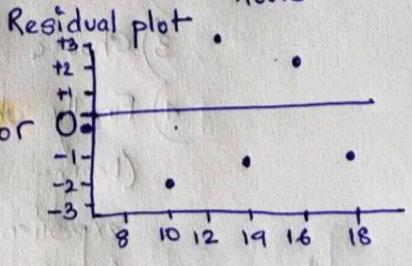
Residual → Error Amount.

In residual plot, x axis is same but y axis changes.

Example –

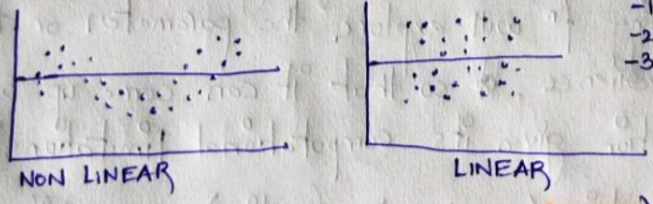
| Hours Studied (X) | Marks Earned (Y) (Actual) | Predicted marks ($\hat{Y}$) (Predicted) | Residual $e = Y - \hat{Y}$ |
|---|---|---|---|
| 10 | 38 | 40 | -2 |
| 12 | 51 | 47.9 | 3.1 |
| 14 | 54 | 55.8 | -1.8 |
| 16 | 66 | 63.7 | 2.3 |
| 18 | 70 | 71.6 | -1.6 |



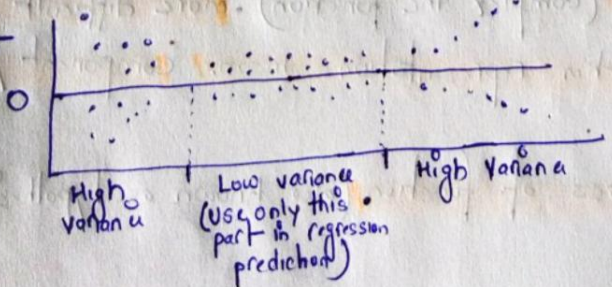




From residual plots we can check

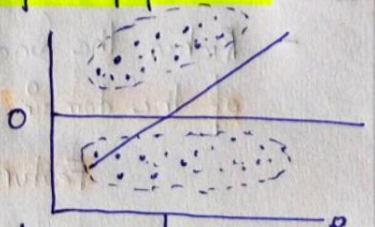
1) Linearity –




2) Variance –




3) Seperate population




Build two seperate regression line for different population.